Estimating a Dutch Value Set for the Pediatric Preference-Based CHU9D Using a Discrete Choice Experiment with Duration

Rowen, Donna; Mulhern, Brendan; Stevens, Katherine; Vermaire, Jan

Published in:
Value in Health

DOI:
10.1016/j.jval.2018.03.016

VALUE IN HEALTH 21 (2018) 1234–1242

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Estimating a Dutch Value Set for the Pediatric Preference-Based CHU9D Using a Discrete Choice Experiment with Duration

Donna Rowen, PhD[1,*], Brendan Mulhern, MSc[2], Katherine Stevens, PhD[1], Jan Hendrik Vermaire, PhD[3,4]

[1]Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, Sheffield, UK; [2]Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW, Australia; [3]TNO Child Health, Oral Health Division, Leiden, The Netherlands; [4]Centrum voor Tandheelkunde en Mondzorgkunde, University of Groningen, Groningen, The Netherlands

ABSTRACT

Objective: This article presents the development of the Dutch value set for the Child Health Utility 9D, a pediatric preference-based measure of quality of life that can be used to generate quality-adjusted life-years. **Methods:** A large online survey was conducted using a discrete choice experiment including a duration attribute with adult members of the Netherlands general population (N = 1276) who were representative in terms of age, gender, marital status, employment, education, and region. Respondents were asked which of two health states they prefer, where each health state was described using the nine dimensions of the Child Health Utility 9D (worried, sad, pain, tired, annoyed, school work/homework, sleep, daily routine, able to join in activities) and duration. The data were modeled using conditional logit with robust standard errors to produce utility values for every health state described by the Child Health Utility 9D. **Results:** The majority of the dimension level coefficients were monotonic, leading to a decrease in utility as severity increases. There was, however, evidence of some logical inconsistencies, particularly for the school work/homework dimension. The value set produced was based on the ordered model and ranges from –0.568 for the worst state to 1 for the best state. **Conclusion:** The valuation of the Child Health Utility 9D using online discrete choice experiment with duration with adult members of the Dutch general population was feasible and produced a valid model for use in cost utility analysis. Normative questions are raised around the valuation of pediatric preference-based measures, including the appropriate perspective for imagining hypothetical pediatric health states.

Keywords: CHU-9D, CHU-9D-NL, discrete choice experiment, pediatric HRQoL, preference-based measures.

Introduction

Economic evaluation of health care interventions often involves the use of incremental cost-effectiveness ratios, where the quality-adjusted life-year (QALY) is used to capture the benefit of different interventions. The QALY is a measure of benefit that captures health impact of conditions and healthcare interventions in terms of its effects on both morbidity and mortality, generated by multiplying a quality adjustment weight by duration to produce a single figure. The quality adjustment weight is often generated using an existing generic preference-based measure such as the EuroQol five-dimensional questionnaire (EQ-5D) [1] or Short-form six-dimension (SF-6D) [2,3]. These have value sets that generate utility values for all health states defined by the measure on the 1–0 full health–dead scale required to generate QALYs. These measures were developed for adults, however, and are not intended for use in children. Currently five pediatric preference-based measures are available. The EQ-5D-Y is a youth version of the EQ-5D intended for use in pediatric populations, but has no available value set to enable it to generate QALYs [4,5]. The Health Utilities Index Mark 2 (HUI2) was originally developed for use in pediatric oncology and included a fertility dimension, but is used mainly as a generic measure of health by assuming fertility is normal [6]. The Assessment of the Quality of Life-6D (AQoL-6D) can be used in pediatric populations and was derived from the adult measure [7]. The 17D is a pediatric measure and the 16D is an adolescent measure, and these were derived from the adult measure, the 15D [8,9]. The Child Health Utility 9D (CHU-9D) is a generic pediatric preference-based measure, that, unlike the other measures, has the advantage that it was specifically developed and worded for use in pediatric populations involving children throughout the development of the classification system [10–12]. Value sets exist for the United Kingdom [13] and Australia [14] enabling the measure to generate QALYs using population-specific value sets for those countries.

* *Address correspondence to:* Donna Rowen, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent St., Sheffield S1 4DA, UK.

E-mail: d.rowen@sheffield.ac.uk

https://doi.org/10.1016/j.jval.2018.03.016

Valuation of pediatric measures is a contentious issue, as there is substantial debate around who should value the measure that captures the health of children. It is a normative question as to whose values should be used to score a pediatric preference-based measure, whether it should be adults or children/adolescents. An argument for the use of child or adolescent values is that this is the group that experiences the health states (though unless they are valuing their own health, the health state will be hypothetical), and the measure is developed for completion by this group, so the values used should reflect that. However, although previous research has suggested that adolescents have understanding of some tasks, such as best–worst scaling and discrete choice experiment (DCE) [14,15], children aged 7 to 11 are unlikely to fully understand any tasks that can reasonably be used to elicit preferences for health states [16]. This raises the question of whether adolescent values are more appropriate for children (aged 7 to 11) than adult values. This is further complicated by the fact that, for ethical reasons, adolescents are usually considered unable to answer questions involving consideration of death, dying, or being dead, meaning that adult (or young adult) values are required to anchor health states on the 1–0 full health–dead scale, for example, through the use of time trade-off or standard gamble (see [17] for an overview of anchoring ordinal data onto the 1–0 scale). The practicalities of obtaining child or adolescent values anchored onto the 1–0 scale and ensuring understanding in children and adolescents participating in health state valuation present considerable challenges.

In contrast, health state valuation of hypothetical health states has been extensively undertaken using adults for a wide range of adult measures, and has been previously used to value pediatric preference-based measures (HUI2, CHU-9D, AQoL-6D). The rationale for using adult values is that because adults typically pay for health care via taxation, and are therefore the funders of the system, it is arguably their preferences that count. This is arguably also consistent with the use of general population values rather than patient values for adult health states. From a pragmatic perspective, adults presumably have a greater understanding of preference elicitation tasks used to elicit preferences for different hypothetical health states and may also be better able to imagine hypothetical health states. In addition, all preference elicitation tasks can be reasonably used in an adult population regardless of whether they mention death, for example, through asking the adult to consider whether he or she would rather be dead than live in a certain health state. However, adult preferences do not necessarily reflect child/adolescent preferences.

This article reports the valuation of the CHU-9D in the Netherlands using online DCE with duration (referred to as DCE_{TTO}, DCE time trade-off) with an adult general population sample, and presents the value set recommended for use to score the measure to generate QALYs for use in economic evaluation. This is a novel application of DCE_{TTO} that has not been used previously to value a pediatric measure. DCE_{TTO} is a relatively new technique that has been successfully used and tested to value several preference-based measures for adults (e.g., [18–23]). Respondents complete a series of choice sets including health states with an associated duration. Responses are modeled to generate a value set anchored on the 1–0 full health–dead scale required to estimate QALYs for all health states described by the classification system. In this article we also compare the new Netherlands value set to the existing CHU-9D value set for the United Kingdom.

Methods

Classification System

The CHU-9D is a pediatric preference-based measure of quality of life suitable for use in children and adolescents aged 7 to 17 years [10–12]. The measure has nine dimensions (worried, sad, pain, tired, annoyed, school work/homework, sleep, daily routine, able to join in activities), each with five severity levels (see Fig. 1). The measure was developed with qualitative interviews with more than 70 school children aged 7 to 11 in the United Kingdom. Thematic content analysis using Framework was used to analyze the data and to select both the dimensions and the wording of the dimensions [11]. The measure has been translated into seven languages including Dutch and has been used in more than 190 studies.

The measure has been valued in the United Kingdom using standard gamble on a representative sample of the adult UK general population where respondents were asked to imagine the hypothetical health state for themselves and were not informed that the health state was a description of pediatric health [13]. The measure has been valued in Australia using a representative sample of adolescents using best–worst scaling [24], where the values were anchored onto the 1–0 full health–dead scale using time trade-off values elicited from a sample of young adults [25]. An equivalent value set also exists using preferences elicited from adults [14].

The Dutch version of the CHU-9D was translated by an ISO 17100-certified translation provider, specialized in patient reported outcome measures (certificate number 3562-TX-0001). The procedure entailed concept elaboration, dual forward translation (including reconciliation), dual back translation (including a review by the CHU-9D developer), cognitive debriefing by five Dutch native speaking residents (7 to 17 years of age; either healthy or with any medical condition), and proofreading by a separate professional linguist.

Preference Elicitation Procedure

Whose values?

This study values the measure using a representative sample of the adult population in the Netherlands as also used in the UK valuation [13]. This was chosen because, first, adults are the taxpayers of the system, and second because the challenges of valuation in young children make adult valuation the most feasible approach for generating considered values.

Valuation technique

Health states have been traditionally valued using techniques such as time trade-off and standard gamble. Time trade-off determines the point at which respondents are indifferent between, say, 10 years in an impaired health state and x years ($x \leq 10$) in full health, where the health state is considered better than being dead. There are, however, well-documented issues with time trade-off and standard gamble techniques including that time trade-off can incorporate time preference and standard gamble can incorporate attitudes to risk, and both typically involve using a different process being to elicit health states worse or better than dead (see [26] for an overview). Recent years have seen increasing usage of online ordinal techniques. Best–worst scaling has been used to value health states [24,27,28], in which respondents are shown a health state with a severity level for each dimension and are typically asked to select the best part and the worst part of the health state. Best–worst scaling cannot produce utility estimates on the 1–0 full health to dead scale without the use of additional preference information about how health states are valued in relation to dead, such as through the use of time trade-off. DCE_{TTO} has been successfully used internationally to value health state classification systems such as the EQ-5D-3L, EQ-5D-5L, and SF-6D [18–23]. DCE_{TTO} has the advantage that it can be successfully used online, allowing for less costly and quicker data collection with no interviewer effect or data inputting errors. In addition, question format does not differ for

Dimension	Severity	Wording
Worry	0	I don't feel worried today
	1	I feel a little bit worried today
	2	I feel a bit worried today
	3	I feel quite worried today
	4	I feel very worried today
Sad	0	I don't feel sad today
	1	I feel a little bit sad today
	2	I feel a bit sad today
	3	I feel quite sad today
	4	I feel very sad today
Pain	0	I don't have any pain today
	1	I have a little bit of pain today
	2	I have a bit of pain today
	3	I have quite a lot of pain today
	4	I have a lot of pain today
Tired	0	I don't feel tired today
	1	I feel a little bit tired today
	2	I feel a bit tired today
	3	I feel quite tired today
	4	I feel very tired today
Annoyed	0	I don't feel annoyed today
	1	I feel a little bit annoyed today
	2	I feel a bit annoyed today
	3	I feel quite annoyed today
	4	I feel very annoyed today
School work/homework	0	I have no problems with my school work/homework today
	1	I have a few problems with my school work/homework today
	2	I have some problems with my school work/homework today
	3	I have many problems with my school work/homework today
	4	I can't do my schoolwork/homework today
Sleep	0	Last night I had no problems sleeping
	1	Last night I had a few problems sleeping
	2	Last night I had some problems sleeping
	3	Last night I had many problems sleeping
	4	Last night I couldn't sleep at all
Daily routine	0	I have no problems with my daily routine today
	1	I have a few problems with my daily routine today
	2	I have some problems with my daily routine today
	3	I have many problems with my daily routine today
	4	I can't do my daily routine today
Able to join in activities	0	I can join in with any activities today
	1	I can join in with most activities today
	2	I can join in with some activities today
	3	I can join in with a few activities today
	4	I can join in with no activities today

Fig. 1 – CHU-9D classification system.

states considered better or worse than dead (unlike many protocols for time trade-off and standard gamble) and (unlike best–worst scaling) no additional data is required to generate estimates anchored on the 1–0 full health to dead scale. Further considerations for states worse than dead are not needed in the DCE_{TTO} task because it is designed to deal with both states better than and worse than dead. This is undertaken through eliciting ordinal preferences, modeling these preferences on a latent scale and then anchoring onto the full health–dead utility scale to determine where dead is placed (described in detail in the Analysis section). This technique is used here, asking respondents to choose "which they prefer" of two options or profiles: health description A for a certain number of years or health description B for a certain number of years (see Figure 2 for an example worded in English). Each health description profile is made up of one severity level of each of the 9 CHU-9D dimensions, alongside the number of years the health state is experienced before death.

Selecting profiles

The CHU-9D classification system generates 1,953,125 health states, meaning that it is infeasible to include all health states in any valuation study. Health states were not selected for valuation per se; instead, choice sets were selected for the DCE, each consisting of two health states with a specified duration. Choice sets were selected to ensure that the collected data would enable the estimation of a prespecified regression model that can be used to generate a value set for all health states defined by the classification system. In the design, CHU-9D health states were paired with one of four duration levels (1y, 4y, 7y, 10y) as successfully used previously to produce logical and valid results [22]. Profiles were selected using D-optimal methods via the experimental design software NGene that takes into account the prespecified regression model to be applied to the data. A modified Fedorov algorithm was used to generate a d-efficient design based on selecting a starting design from a candidate set and iteratively improving that design to minimize the d-error. The design was selected when the d-error of the design was not improved after 2 minutes of further iteration. No prior values were used given that the CHU-9D has not previously been valued using the DCE_{TTO} method internationally, and no Dutch value set exists using another valuation method. There are 10 dimensions in each profile, which is a great deal of information for respondents to simultaneously consider and process when choosing between two profiles. To simplify the task, the design imposed

Health description A	Health description B
You live for 10 years with the following then you die:	You live for 1 year with the following then you die:
You feel a little bit worried	You feel a little bit worried
You feel a bit sad	You feel very sad
You have a bit of pain	You don't have any pain
You feel quite tired	You feel quite tired
You feel quite annoyed	You don't feel annoyed
You can't do work/housework	You have many problems with your work/housework
You have a few problems sleeping	You can't sleep at all
You can't do your daily routine	You have a few problems with your daily routine
You can join in with any activities	You can join in with any activities

Which do you prefer?

Fig. 2 – Example discrete choice experiment (DCE) task.

a constraint that the severity level of three of the nine health dimensions were fixed in a given choice set, an approach that has been successfully used in a previous study [23]. In total, 204 choice sets were selected across 17 survey versions, with 12 choice sets per survey. Across all choice sets and survey versions 408 different health states were included. The design also allocated profiles to either A or B (right or left of the screen) and allocated choice sets to survey versions. There is no set guidance regarding the sample size required for each choice set to estimate the model parameters with confidence. Nevertheless, the sample size of 1200 aimed for in this study, completing 12 tasks each, results in approximately 70 observations per choice set, which is in the range of other (referred to as DCE_{TTO}, DCE time trade-off) studies. Choice sets were randomly ordered within each survey version for each respondent, but all choice sets had the same order of dimensions.

Respondents

Respondents were recruited from an existing online panel of respondents of the adult general population in the Netherlands by a market research agency (KIEN Research, Groningen, The Netherlands). Potential respondents were requested to participate in the survey by email, and were sampled to be representative of the adult Netherlands population in terms of age, gender, marital status, education, employment status, and region. After completion of the survey, respondents received a nominal reward for their participation from the market research agency, in line with that for other online panel-based surveys.

Perspective

It is a normative question as to which perspective should be used for the valuation of a pediatric measure. For example, respondents could be asked to imagine that they are in the health state as themselves (i.e., as an adult), or as a child, or to imagine a child is in the health state. This study asked respondents which health state they prefer, after imagining that they themselves are in each state, as also used in the UK valuation [13]. Respondents were not informed that the health state was pediatric, and hence the school work/homework dimension was reworded to work/housework to be inclusive of all respondents in the sample regardless of whether they were working or not.

The DCE survey

Sampled respondents were contacted by email from the market research agency requesting their participation in the survey. The DCE_{TTO} task was designed for different platforms, and could be answered on desktop and laptop computers as well as mobile devices including tablets and smartphones. The survey began with the respondent reading an information sheet about the project and giving informed consent. Respondents then answered questions on their sociodemographic characteristics and their health, including on the EQ-5D-5L. Respondents were then asked to complete the CHU-9D for themselves, where the school work/homework dimension was reworded to work/housework, and this was undertaken to familiarize respondents with the classification system. Respondents then answered a warm-up practice DCE question (see Figure 2 for an example worded in English) and 12 DCE tasks. Dimensions that were identical within a choice set were indicated using light gray text, and dimensions that differed within the choice set had black text. Finally, respondents were asked how difficult they found it to choose between the different health descriptions.

Ethical approval was granted via the Central Committee on Research Involving Human Subjects CCMO NL201623-6, which concluded this research did not fall under the Law of Medical Research Involving Human Subjects (WMO law).

Analysis

Summaries of sociodemographic and health characteristics of the sample were generated and compared to the adult general population of the Netherlands. The DCE with duration data was analyzed using the model specified in [18]:

$$\mu_{ij} = \alpha_i + \beta_1 t_{ij} + \beta'_2 \mathbf{x}_{ij} t_{ij} + \varepsilon_{ij} \quad (1)$$

where μ_{ij} represents the utility of individual i for health state profile j,$\alpha_i$is an individual specific constant term, ε_{ij} represents the error term, β_1 is the coefficient for duration in life years t, and β'_2 represents the coefficients on the 36 interaction terms of duration and attribute variables composed of levels 1, 2, 3, and 4 of each quality of life attribute (where level 0 is the baseline). Duration was modeled as a linear, continuous variable. This was examined by modeling duration as a categorical variable and plotting the duration coefficients [29]. Severity levels of each dimension ($\mathbf{x}_{ij}$) were not entered without being interacted with duration, as a health state cannot be meaningfully valued without consideration of its duration, and inclusion of both $\mathbf{x}_{ij}$ and $\mathbf{x}_{ij} t_{ij}$ would suffer from multicollinearity. The data were modeled using conditional logit with robust standard errors.

The specification in Eq. (1) generates utility values on a latent utility scale, which cannot be easily interpreted. The latent values are anchored (called "anchored values" and "value set" in the text that follows) onto the 1–0 full health-dead utility scale required to generate QALYs using the marginal rate of substitution, where the coefficient for each level of each dimension is divided by the coefficient for duration to generate a utility weight for a given severity level of a dimension: $\frac{\beta_{2ij}}{\beta_1}$. Utility values for health states are generated as 1 plus the sum of the utility weights for the relevant severity level of each dimension.

Table 1 – Characteristics of the sample providing valuation data for the Dutch valuation of the CHU-9D.

Characteristic	Sample, % (N = 1276)	Netherlands population % (N = 16,979,120)
Female	50.2	50.2
Age group		
Younger than 30	16.5	18.7
30–39	15.4	15.3
40–49	18.8	18.9
50–59	17.0	17.7
60+	32.2	29.4
Highest education*		
Low	28.5	31.7
Middle	45.5	43.7
High	25.9	24.6
Household		
Live alone	21.9	21.4
Household without children	51.1	49.1
Household with children (youngest aged 12 or younger)	19.8	21.7
Household with children (youngest aged 13–17)	7.2	7.8
Region		
North Netherlands	10.6	10.1
East Netherlands	20.5	21.1
South Netherlands	25.1	21.3
West Netherlands	43.8	47.5
Employment status		
Full time	34.7	31.7
Part time	21.2	21.9
Not working	44.0	46.5
Home ownership		
Owner occupied	59.6	57.5
Rent	37.5	42.5
Other	2.8	—
Marital status	%	
Married/cohabiting	63.8	62.3
Single (divorced)	8.4	7.5
Single (not separated)	13.4	17.7
Widow/widower	4.7	5.0
In a relationship but not living together	7.6	7.5
Other	2.1	—
EQ-5D-5L, mean (SD)	0.795 (0.230)	0.869 (0.170)

* Education levels: low—preparatory secondary vocational education or lower; middle—senior secondary vocational education, senior secondary general education; high—higher professional education or higher.

To inform policy a logically consistent model is required, where the utility value either remains the same or decreases as health and quality of life deteriorate. For example, if a dimension worsens in severity from level 0 to level 1 (e.g., from "I don't feel worried today" to "I feel a little bit worried today," the utility value of the overall health state must not increase. To produce a fully consistent model, adjacent inconsistent levels are merged together; for example, if levels 1 and 2 are inconsistent in a given dimension they can be merged to produce a single utility decrement that is applied if the dimension is at level 1 or level 2. This approach has been widely used previously to produce fully consistent models for use to inform policy (e.g., see [3,13,30–33]).

Results

The Sample

In total, 1276 respondents representative of the Dutch population in terms of age and gender fully completed the survey (see Table 1). The sample is nationally representative for highest level of education, household composition (in terms of living alone, or living with/without children), employment status, home ownership, marital status, and regional differences (rural vs. urban). EQ-5D-5L (scored using [34]) is lower in the sample than in the population norms, with mean 0.795 (SD 0.230) in comparison to the population norm 0.869 (SD 0.170). Worse health in comparison to population norms has also been reported in other large online surveys using participants recruited by a market research agency (e.g., see [20,35]).

A total of 12.9% respondents found it very difficult to make a choice between the health states in the DCE tasks, 42.9% found it difficult, 32.8% found it neither difficult nor easy, 9.7% found it easy, and 1.7% of the participants found it very easy to make a choice.

Regression Analysis

The first model estimated included all data and imposed no restrictions on the model in terms of logical consistencies, where utility decrements increase as severity worsens within any given dimension (see Table 2). Duration is significant and positive as expected, where health states with longer duration are preferred to health states with shorter duration. Prior analyses (not reported) indicated it was appropriate to assume that duration was linear and continuous, through the use of a plot of the duration levels and coefficients estimated using a model in which duration was included as a categorical variable.

The majority of interaction terms of dimension levels multiplied by duration are significant, have the expected sign, and are logically consistent. Coefficients for sad level 1*duration, annoyed level 1*duration, and work/housework level 1*duration have the wrong sign (in comparison to the baseline level 0), meaning they are logically inconsistent. For sad level 1*duration and annoyed level 1*duration the coefficients are nonsignificant and for work/housework level 1*duration the coefficient is significant at the 5% level but all coefficients are small, meaning that for these dimensions respondents have little distinction in terms of the impact on their utility between levels 0 and 1.

Inconsistent and significant coefficients are observed for adjacent levels of pain level 1*duration and pain level 2*duration, tired level 1*duration and tired level 2*duration, and work/housework level 3*duration and work/housework level 4*duration.

The consistent model (Table 2) removes all logical inconsistencies and merges sad level 1*duration, annoyed level 1*duration and work level 1*duration with the reference level for these dimensions. Pain level 1*duration is merged with pain level 2*duration, tired level 1*duration is merged with pain level 2*duration, and work/housework level 3*duration is merged with work/housework level 4*duration. All coefficients are significant with the exception of worried level 1*duration, annoyed level 2*duration, and able to join in activities level 1*duration. The utility values range from 1 for the best state to –0.568 for the worst state.

Table 2 – Regression analysis of DCE data, all data included, estimates on a latent utility scale.

	First model		Consistent model
Worry level 1*duration	–0.013	Worried level 1*duration	–0.013
Worry level 2*duration	–0.050***	Worried level 2*duration	–0.050***
Worry level 3*duration	–0.055***	Worried level 3*duration	–0.054***
Worry level 4*duration	–0.087***	Worried level 4*duration	–0.087***
Sad level 1*duration	0.001		
Sad level 2*duration	–0.022***	Sad level 2*duration	–0.023***
Sad level 3*duration	–0.055***	Sad level 3*duration	–0.057***
Sad level 4*duration	–0.075***	Sad level 4*duration	–0.074***
Pain level 1*duration	–0.036***	Pain level 1 or 2*duration	–0.033***
Pain level 2*duration	–0.031***		
Pain level 3*duration	–0.091***	Pain level 3*duration	–0.091***
Pain level 4*duration	–0.145***	Pain level 4*duration	–0.145***
Tired level 1*duration	–0.027***	Tired level 1 or 2*duration	–0.023***
Tired level 2*duration	–0.026***		
Tired level 3*duration	–0.045***	Tired level 3*duration	–0.042***
Tired level 4*duration	–0.072***	Tired level 4*duration	–0.070***
Annoyed level 1*duration	0.012		
Annoyed level 2*duration	–0.003	Annoyed level 2*duration	–0.008
Annoyed level 3*duration	–0.041***	Annoyed level 3*duration	–0.046***
Annoyed level 4*duration	–0.051***	Annoyed level 4*duration	–0.058***
Work/housework level 1*duration	0.017**		
Work/housework level 2*duration	–0.011	Work/housework level 2*duration	–0.018**
Work/housework level 3*duration	–0.047***	Work/housework level 3 or 4*duration	–0.056***
Work/housework level 4*duration	–0.047***		
Sleep level 1*duration	–0.018**	Sleep level 1*duration	–0.019**
Sleep level 2*duration	–0.033***	Sleep level 2*duration	–0.033***
Sleep level 3*duration	–0.068***	Sleep level 3*duration	–0.070***
Sleep level 4*duration	–0.118***	Sleep level 4*duration	–0.118***
Daily routine level 1* duration	–0.017**	Daily routine level 1* duration	–0.018**
Daily routine level 2* duration	–0.022***	Daily routine level 2* duration	–0.022***
Daily routine level 3* duration	–0.079***	Daily routine level 3* duration	–0.079***
Daily routine level 4* duration	–0.097***	Daily routine level 4* duration	–0.096***
Able to join in activities level 1*duration	–0.001	Able to join in activities level 1*duration	–0.003
Able to join in activities level 2*duration	–0.030***	Able to join in activities level 2*duration	–0.029***
Able to join in activities level 3*duration	–0.040***	Able to join in activities level 3*duration	–0.041***
Able to join in activities level 4*duration	–0.101***	Able to join in activities level 4*duration	–0.099***
Duration	0.500***	Duration	0.513***
Observations	30,474	Observations	30,474
Log likelihood	–8271	Log likelihood	–8274
Rho-squared	0.217	Rho–squared	0.217

Notes. Significance is reported for the interaction terms.*** $P < 0.01$, **$P < 0.05$, *$P < 0.1$.

Dutch Value Set

Table 3 presents the Dutch value set, generated using the anchored values of the consistent model. Utility weights were generated using the marginal rate of substitution as outlined in the Methods section. The utility value for a health states is 1 plus the sum of the utility weights for each relevant severity level of each dimension. For example, state 401200300 is generated using $1 + (–0.170 + 0 – 0.065 – 0.045 + 0 + 0 – 0.136 + 0 + 0) = 0.584$.

Comparison to Existing UK Value Set

Figure 3 shows a comparison of the Dutch value set and the existing UK value set. The worst state (–0.568) is substantially lower than the UK value (0.326). The utility weights are larger for the Dutch value set in comparison to the UK value set.

The UK consistent model involved the merging of several adjacent severity levels, meaning that for the dimensions of worry, tired, and annoyed in particular there is the same utility decrement for any deterioration in severity from level 0, meaning that for these dimensions there is no change in utility resulting from any change within the levels 1, 2, 3, or 4. In contrast, in the Dutch value set for the worry dimension the utility decrements vary from –0.025 for level 1 to –0.170 for level 4, for the tired dimension the utility decrements vary from –0.045 for level 1 to –0.136 for level 4, and for the annoyed dimension the utility decrements vary from 0 for level 1 to –0.113 for level 4. The work/housework dimension in the Netherlands value set is the only dimension in which more than two levels are merged (levels 0 and 1, levels 3 and 4), and this was similar to the UK value set, where the work dimension was merged for levels 1 and 2, and merged for levels 3 and 4.

Discussion

This article has presented the valuation of the CHU-9D using online DCE_{TTO} with a nationally representative sample of the adult general population in the Netherlands. The approach was feasible and generated sensible results. The first estimated model had six logical inconsistencies for the dimensions of sad, pain,

Table 3 – Dutch value set for the CHU-9D (utility estimates are anchored on the 1–0 full health–dead scale required to generate QALYs).

Dimension	Severity	Utility decrement	Standard error
Worry	0	0	
	1	−0.025	0.015
	2	−0.097	0.013
	3	−0.106	0.016
	4	−0.170	0.014
Sad	0	0	
	1	0	
	2	−0.045	0.013
	3	−0.111	0.014
	4	−0.145	0.014
Pain	0	−0	
	1	−0.065	0.014
	2	−0.065	0.014
	3	−0.177	0.015
	4	−0.283	0.017
Tired	0	0	
	1	−0.045	0.014
	2	−0.045	0.014
	3	−0.082	0.016
	4	−0.136	0.015
Annoyed	0	0	
	1	0	
	2	−0.016	0.014
	3	−0.090	0.014
	4	−0.113	0.013
School work/ homework	0	0	
	1	0	
	2	−0.034	0.013
	3	−0.110	0.011
	4	−0.110	0.011
Sleep	0	0	
	1	−0.037	0.016
	2	−0.063	0.016
	3	−0.136	0.014
	4	−0.231	0.015
Daily routine	0	0	
	1	−0.035	0.015
	2	−0.042	0.014
	3	−0.155	0.017
	4	−0.186	0.015
Able to join in activities	0	0	
	1	−0.005	0.016
	2	−0.057	0.016
	3	−0.079	0.018
	4	−0.194	0.016

tired, annoyed, and work/housework multiplied by duration, meaning that a worsening in severity between certain severity levels led to an increase in utility. For use in policy a logically consistent model is required, where utility decreases or remains the same as severity increases, and hence a consistent model was estimated. This is the recommended value set for use in the Netherlands for the economic evaluation of health interventions for pediatric populations.

The most inconsistent dimension was work/housework. This dimension was reworded from the dimension in the CHU-9D classification system that refers to school work/homework to ensure it was relevant to adults valuing the health state from their own perspective. Children reporting their quality of life using the CHU-9D will respond to the school work/homework dimension by reporting problems with school work/homework, whereas adults are valuing problems with their work/housework. It is likely that the rewording of the dimension to work//housework from school work/homework has changed the meaning for this dimension. For example, the utility impact of problems with work/housework for adults may include income effects and impact on others including coworkers and family/friends needing to compensate, which would not be expected to be the case with children, who would need to catch up on any school work/homework missed once they are able to or risk falling behind relative to their peers. There is a large contrast in the size of the coefficients in the Dutch value set in comparison to the UK value set. This has been found in the literature, and many measures such as the EQ-5D and the SF-6D have different value sets for many different countries, as different countries have different preferences for health (e.g., see [36]). This is likely due to many factors, including cultural, social, and work differences; however, differences in elicitation techniques and study protocols can also be a contributing factor. Sociodemographic characteristics of the selected sample can also have an effect on on utility values (e.g., see [37]), and as different countries have different sociodemographic profiles this may also have an effect on the values.

It is also likely, however, that the difference between the value sets is at least in part due to the different elicitation techniques used in the UK and Dutch value sets. Both the UK and the Dutch studies used the same perspective, yet the UK value set used standard gamble administered in face-to-face interviews whereas the Dutch value set used online DCE with a duration attribute. Some value set differences would be expected owing to cultural and work differences but some differences may be due to the different elicitation techniques. Values generated using standard gamble can be impacted on by respondents' attitudes to risk, meaning that values can be relatively high for severe health states. In contrast, DCE_{TTO} instead asks people (implicitly) to trade length of life with quality of life. The version of DCE_{TTO} used here involved a forced choice, where respondents could not say that they had an equal preference for each health description, whereas in standard gamble respondents can state indifference. Standard gamble has been found to produce higher values than time trade-off (TTO) [38], and TTO has been found to produce higher values than DCE_{TTO} [18]. This means that the finding here that standard gamble produced higher values than DCE_{TTO} may have been expected. This has also been found previously when comparing UK SF-6D values elicited using standard gamble to Australian SF-6D values elicited using DCE_{TTO} [22]. Further research comparing DCE_{TTO} to other elicitation techniques is encouraged.

The differences between the UK and Dutch value sets may have policy implications. Comparing the Netherlands value set to the UK value set shows a contrast in the number of merged adjacent severity levels in the two value sets. This means that it is expected that the Dutch value set will be more responsive in terms of a change in utility when quality of life changes are reported. Further research is being conducted to determine the effect of using the Dutch values in comparison to the UK values both in terms of policy implications and the psychometric properties of the measure.

Limitations of the study include the use of an online survey, where respondent engagement and understanding cannot be accurately measured. This may be particularly relevant for this study, which included a large number of attributes in the DCE tasks, which may have been cognitively challenging even with overlap of dimensions built into the study design. Nearly 13% of respondents stated that they found it very difficult to choose

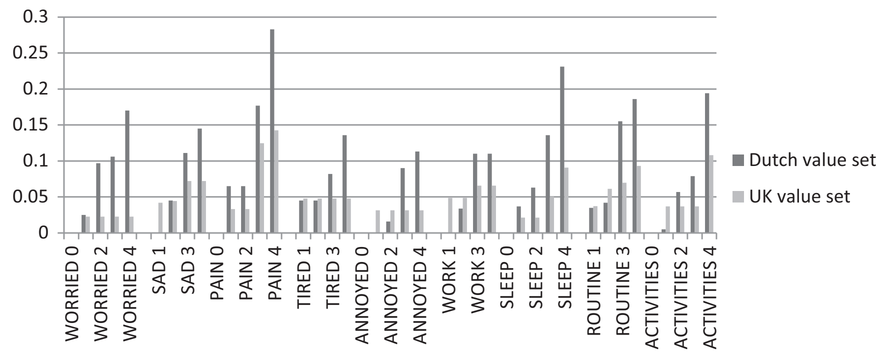

Fig. 3 – Plot of coefficients comparing Dutch value set to the existing UK value set.

between the different health descriptions, and nearly 56% of respondents stated that they found it difficult. These results are likely to reflect both that respondents found it difficult to choose which health description they preferred and that they found the task cognitively demanding. As the study aim was to produce a value set representative of the population preferences, the data have not been modeled to take into account preference heterogeneity, where preferences may vary according to observable or unobservable characteristics of respondents, and this may have affected the results.

Choice of Perspective

The choices made in this study regarding perspective, where a group of adults were asked to value hypothetical health states, imagining that they were experiencing the health states and not being informed they were descriptions of pediatric health, is contentious. However, there is no agreed protocol for the valuation of measures assessing child health, and the study design choices made involved a normative decision. Other alternatives include asking adults to provide values using different perspectives including “imagining you are a 10-year-old child” or “considering your views about a 10-year-old child” [6,39]. The impact of taking these different approaches is currently being explored by The EuroQol Group in the context of the valuation of the EQ-5D-Y.

One concern with the use of the perspective where adults imagine themselves as a child is that the values they provide may be affected by problems of recall, as they may not accurately be able to imagine themselves or their true preferences as a child. Taking the perspective in which adults consider their views about a 10–year-old child will lead to values being influenced by the individual’s experience of children, including whether or not the respondent has children, and their level of exposure to children both in the past and present. The values provided will be impacted by which 10-year-old child the respondent is considering, for example, their son or daughter, their grandchild, their friend’s child, and so forth. It is likely that this would mean that the elicited utility values would not simply reflect the perceived impact of the health state on the child, but also the respondent’s perception of how bad they think in general it is for a child, and for the child they are thinking of in particular, to have any health problem. This is supported in the literature, where one study found that Visual Analog Scale (VAS) values were lower when respondents were asked to imagine another adult in comparison to imagining they applied to a 10-year-old child [39]. There is also the possibility that respondents are not willing to trade between years of life and quality of life for a child in the same way that they are prepared to trade between years of their own life and their own quality of life, and this possibility is also being explored by The EuroQol Group in the context of the valuation of EQ-5D-Y. If it is found that respondents are not willing to sacrifice years of a child’s life in exchange for improving their quality of life, then this also has an effect on the choice of elicitation technique. For example, DCE_{TTO} that is used in this study involves trading between years of life and quality of life, and hence is unlikely to be appropriate under these circumstances.

If there is something in particular about what children or adolescents experience themselves when they are sick that makes utility for children or adolescents different from that for adults, then arguably an adult cannot accurately imagine this without further information. The use of “informed” adult values could offer a solution in which respondents are provided with details on what patients experience and how they feel about how different dimensions affect them (see [40] for an overview). In turn, child or adolescent valuation is also an option, as it can be argued that children and adolescents should be asked to value the health states as children and adolescents experience the health states and better understand the effect of the quality of life problems on their life. For the reasons outlined in the introduction, however, this requires further research because of the considerable challenges around the elicitation and anchoring process used to generate utilities on the full-health to dead scale [15,25]. This article has elicited adult values for pediatric states, yet there is no claim that these represent child or adolescent values, as it may be expected that children and adolescents would place a different relative value on different dimensions and severity levels within these. This implies that the use of child values or adult values will potentially have an effect on incremental cost-effectiveness ratios and potentially on resource allocation decisions, meaning that the choice of whose values to use is of extreme importance. One advantage of eliciting adult values is that the adult general population consists of the taxpayers funding the health care system, and it is arguably their preferences that count. This argument is consistent with selecting members of the adult general population to value hypothetical adult health states rather than patients. This consistency in the elicitation of values from the adult general population for hypothetical health states—pediatric, adolescent, and adult—is a considerable advantage for use in an economic model that combines utilities to generate QALYs over a patient’s lifetime from birth to adulthood and beyond.

Finally, the choice of perspective should not be impacted by whether society thinks we should care more or differently about health problems experienced by children. Arguably this should instead be accounted for at the policy level, for example through the use of QALY weightings (where children could be given a higher weight) or a higher cost-effectiveness threshold, meaning more expensive treatments are recommended because they are used to treat child health problems.

This article presented the Dutch value set for the CHU-9D, and is recommended for scoring for use in economic evaluation and other assessments of quality of life in the Netherlands. Further work should assess the psychometric performance of the utility value set in different health conditions to increase confidence in its use. The choice of population, perspective, and technique used to value pediatric preference-based measures is a contentious issue, and further research determining the impact of these choices is encouraged.

Acknowledgments

We would like to acknowledge all the participants in the surveys.

REFERENCES

[1] Brooks R, The EuroQol Group. EuroQol: the current state of play. Health Policy 1996;37:53–72.
[2] Brazier J, Roberts J, Deverill M. The estimation a preference-based single index measure for health from the SF-36. J Health Econ 2002;21:271–92.
[3] Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42:851–9.
[4] Ravens-Sieberer U, Wille N, Badia X, et al. Feasibilty, reliability and validity of the EQ-5D-Y: results from a multinational study. Qual Life Res 2010;19:887–97.
[5] Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res 2010;19:875–86.
[6] Torrance G, Feeny D, Furling W, et al. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index mark 2. Med Care 1996;34:702–22.
[7] Moodie M, Richardson J, Rankin B, et al. Predicting time trade-off health state valuations of adolescents in four pacific countries using the AQoL-6D instrument. Value Health 2010;13:1014–27.
[8] Apajasalo M, Sintonen H, Holmberg C, et al. Quality of life in early adolescence: a sixteen-dimensional health-related measure (16D). Qual Life Res 1996;5:205–11.
[9] Apajasalo M, Rautonen J, Holmberg C, et al. Quality of life in pre-adolescence: a 17-dimensional health-related measure (17D). Qual Life Res 1996;5:532–8.
[10] Stevens KJ. Developing a descriptive system for a new preference-based measure of health-related quality of life for children. Qual Life Res 2009;18:1105–13.
[11] Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric health-related quality-of-life measure. Qual Health Res 2010;20:340–51.
[12] Stevens KJ. Assessing the performance of a new generic measure of health related quality of life for children and refining it for use in health state valuation. Appl Health Econ Health Policy 2011;9:157–69.
[13] Stevens K. Valuation of the Child Health Utility 9D Index. Pharmacoeconomics 2012;30:729–47.
[14] Ratcliffe J, Huynh E, Stevens K, et al. Nothing about us without us? A comparison of adolescent and adult health-state values for the child health utility-9D using profile case best-worst scaling. Health Econ 2016;25:486–96.
[15] Ratcliffe J, Couzner L, Flynn T, et al. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare Best-Worst Discrete Choice Experiment, Standard Gamble and Time Trade Off methods. Appl Health Econ Health Policy 2011; 9:15–27.
[16] Stevens KJ. "Because that's what matters to me." A pilot study to test the feasibility and reliability of ordinal valuation methods for health state valuation with children. HEDS Discussion Paper. 2015. Available from: https://www.shef.ac.uk/scharr/sections/heds/discussion-papers/15-05-1.526948. [Accessed February 13, 2018].
[17] Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health–dead QALY scale. Medical Decis Making 2015;35328–40.
[18] Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ 2012;31:306–18.
[19] Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. Soc Sci Med 2014;114:38–48.
[20] Mulhern B, Bansback N, Brazier J, et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. Health Technol Assess 2014;18:vii–xxvi, 1–191.
[21] Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl Health Econ Health Policy 2013;11:287–98.
[22] Norman R, Viney R, Brazier J, et al. Valuing SF-6D health states using a discrete choice experiment. Med Decis Making 2014;34:773–86.
[23] Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Qual Life Res 2016;25:637–49.
[24] Ratcliffe J, Flynn T, Terlich F, et al. Developing adolescent specific health state values for economic evaluation: an application of profile case best worst scaling to the Child Health Utility-9D. Pharmacoeconomics 2012;30:713–27.
[25] Ratcliffe J, Chen G, Stevens K, et al. Valuing Child Health Utility 9D health states with young adults: insights from a time trade off study. Appl Health Econ Health Policy 2015;13:485–92.
[26] Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press, 2016.
[27] Coast J, Flynn T, Natarajan L, et al. Valuing the ICECAP capability index for older people. Soc Sci Med 2008;67:874–82.
[28] Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A capability instrument: estimation of a UK general population tariff. Health Econ 24:258–69.
[29] Payne K, Fargher EA, Roberts SA, et al. Valuing pharmacogenetic testing services: a comparison of patients' and health care professionals' preferences. Value Health 2011;14:121–34.
[30] Mukuria C, Rowen D, Brazier JE, et al. Deriving a preference-based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. Value Health 2015;18:846–55.
[31] Mulhern B, Rowen D, Jacoby A, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. Epilepsy Behav 2012; 24:36–43.
[32] Rowen D, Mulhern B, Banerjee S, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-proxy. Value Health 2012;15:346–56.
[33] Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value Health 2011;14:721–31.
[34] Versteegh MM, Vermeulen KM, Evers SMMA, et al. Dutch tariff for the five-level version of EQ-5D. Value Health 2016;19:343–52.
[35] Rowen D, Brazier J, Mukuria C, et al. Eliciting societal preferences for weighting QALYs for burden of illness and end of life. Med Decis Making 2016;36:210–22.
[36] Ferreira LN, Ferreira PL, Rowen D, Brazier JE. Do Portuguese and UK health state values differ across valuation methods? Qual Life Res 2011;20:609–19.
[37] Dolan P, Roberts J. To what extent can we explain time trade-off values from other information about respondents? Soc Sci Med 2002;54:919–29.
[38] Lenert LA, Cher DJ, Goldstein MK, et al. The effect of search procedures on utility elicitations. Med Decis Making 1998;18:76–83.
[39] Kind P, Klose K, Gusi N, et al. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. Qual Life Res 2015;24:2519–39.
[40] McTaggart-Cowan H. Elicitation of informed general population health state utility values: a review of the literature. Value Health 2011;14:1153–7.